Question 2:

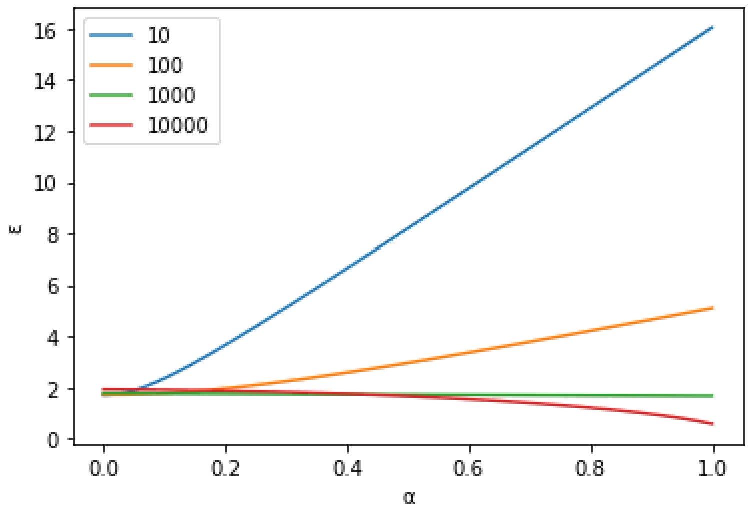
(a)

	α=0.1	α=0.5	α=0.9
(i) N_T=1, N_s=100	5.86	21.61	38.64
(ii) N_T=10, N_s=1000	2.25	8.12	14.46
(iii) N_T=100, N_s=10000	0.86	2.94	5.19
(iv) N_T=1000, N_s=100000	0.36	1.06	1.82

(v)

We notice that as we increase the importance of error in the target domain α for a given setting, the cross-generalization error bound increases because β is low (the number of labeled source domain data is much higher than the target domain). When N_T=1000, N_s=100000 and α=0.1, we get a generalization bound of 0.36 which is less than 0.5, which might be expected as we get a lower generalization error when we increase the number of training points and lower importance of target domain data.

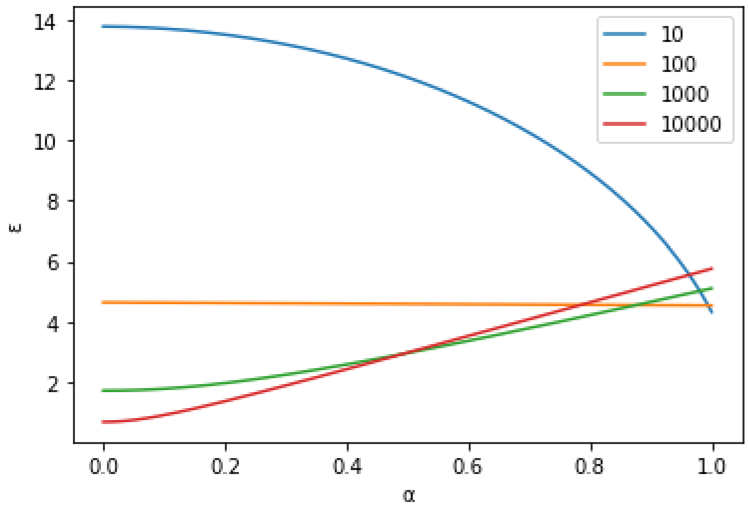
(b)



For N_T=10, optimal α= 0.000629 ≈0
For N_T=100, optimal α= 0.006888≈0
For N_T=1000, optimal α= 0.99999≈1
For N_T=10000, optimal α= 0.99999≈1

As seen in the plots, for lowers values of target domain training points (small β) , ε increases as α increases because the dominant term in equation 6 will be α^2/β which is directly proportional to ε , so we got small optimal values of α for the first 2 cases . On the other hand, for higher values of N_T, when it exceeds the number of points in the source domain N_S, β will be relatively large and the term $(1-\alpha)^2/(1-\beta)$ will be more dominant and it has inverse proportion with ε , and hence ε decreases as α increases for $0<\alpha<1$. Thus, we got large optimal α values for the last two cases.

(c)



For N_s=10, optimal α= 0.999999≈1
For N_s=100, optimal α= 0.999999≈1
For N_s=1000, optimal α= 0.006888≈0
For N_s=10000, optimal α= 0.00175≈0

Following a similar argument, we can see in eq.6 that for lower values of N_S, β will be high which makes $(1-\alpha)^2/(1-\beta)$ is the dominant term and hence ε decreases with α, so we got high optimal values of α for the first two cases. For higher values of N_S, β will be relatively small which makes the term α^2/β more dominant and hence ε increases with α, so we got low optimal values of α for the last two cases.

(d)

(i)

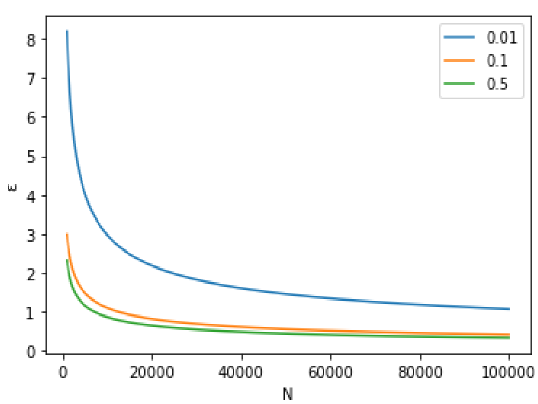
default choice $\alpha = \beta$ is better since it gives lower cross_domain generalization bound ε for all cases from part (b) and (c), as seen in the following tables:

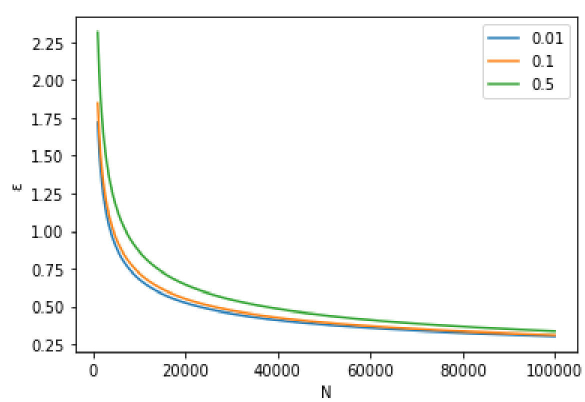
	$\alpha = 0.5$	$\alpha = \beta$
N_T=10, N_s=1000	8.194468756126083	1.7117759913120953
N_T=100, N_s=1000	2.9587345942832277	1.76331389579287
N_T=1000, N_s=1000	1.7216700323794591	1.7216700323794591
N_T=10000, N_s=1000	1.6588500994779338	0.9341146778434084
N_T=100, N_s=10	12.085523466988201	6.9290303614707405
N_T=100, N_s=100	4.587719163486929	4.587719163486929
N_T=100, N_s=1000	2.9587345942832277	1.76331389579287
N_T=100, N_s=10000	2.970587850198958	0.6773441287522688

Which is expected since all plots are increasing or decreasing, and $\alpha = \beta$ will at the lower side of the plot rather than in the middle, and hence ε will be lower at $\alpha = \beta$.

(iii)

$\alpha = 0.5$ $\alpha = \beta$





As expected and as seen in equation 6, N and ε have inverse relation which is evident in the above plots. Also, we have lower generalization bounds for $\alpha = \beta$ than for $\alpha = 0.5$ in the three settings, which is expected since we showed that $\alpha = \beta$ have lower generalization error bounds both algebraically and experimentally in the previous parts.